DOCUMENT RESUME

ED 142 567 TM 006 359

AUTHOR	Ebel, Robert L.
TITLE	The Uses of Standardized Testing. Fastback Series No. 93.
INSTITUTION	Phi Delta Kappa Educational Foundation, Bloomington, Ind.
PUB DATE	77
NOTE	51p.
AVAILABLE FROM	Phi Delta Kappa, Eighth and Union, Box 789, Bloomington, Indiana 47401 (for members, $0.60 ea., for nonmembers, $0.75)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	Academic Achievement; *Achievement Tests; Criterion Referenced Tests; *Educational Assessment; Effective Teaching; Elementary Secondary Education; Formative Evaluation; Norm Referenced Tests; Objective Tests; *Standardized Tests; *Student Evaluation; Summative Evaluation; Test Bias; Testing Problems; Test Reliability; Test Results; Test Validity
IDENTIFIERS	Alternatives to Standardized Testing

ABSTRACT

In a discussion of the value of standardized tests and their implications for effective teaching, the author contends that such tests are useful in keeping those concerned with the learning process focused on the main target. In addition, standardized tests serve to (1) reveal variations in the effectiveness of different instructional procedures or curricula, (2) motivate teachers and students to succeed, (3) recognize and reward success in learning, (4) facilitate comparisons of the effectiveness of students, teachers, and schools, and (5) focus on schooling as a purposeful educational activity with results that can be assessed and documented systematically. The limitations of standardized testing and alternatives to testing are discussed. (MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

Fastback 93

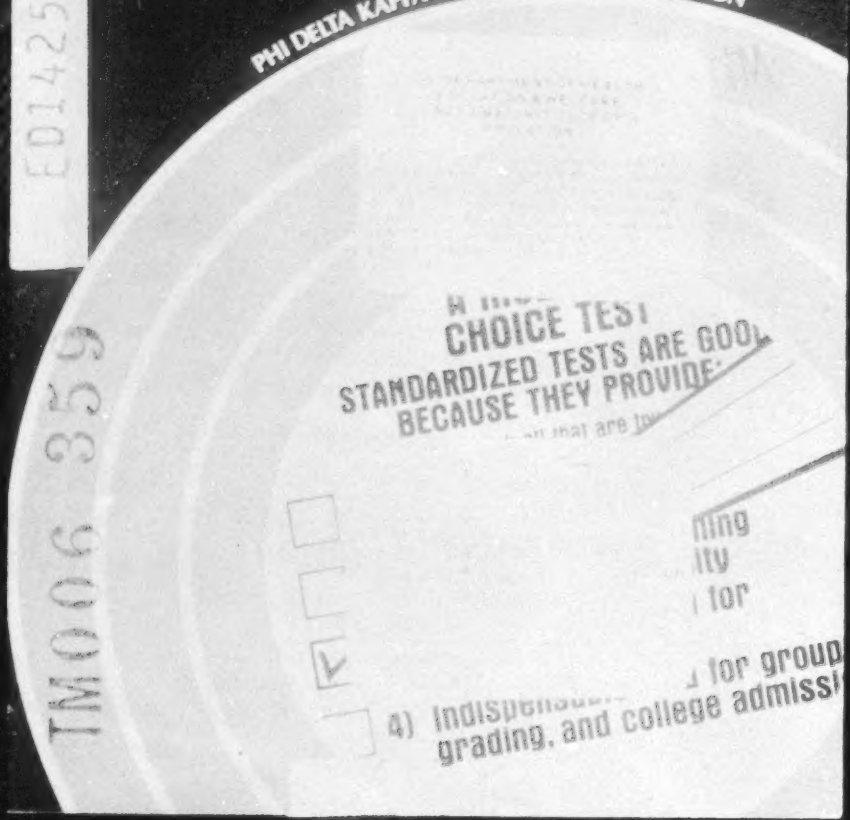
The Uses of Standardized Testing

Robert L. Ebel

ED142567

PHI DELTA KAPPA EDUCATIONAL FOUNDATION

TM006 359



ROBERT L. EBEL

Robert L. Ebel, professor of education and psychology at Michigan State University, is a specialist in educational measurement and research. He joined the Michigan State faculty in 1963 after serving as a vice president of the Educational Testing Services, Princeton, New Jersey, for six years. At ETS he was responsible for test development, statistical analysis, and the direction of testing programs.

Before joining the ETS staff, Ebel was professor of education and director of the Examination Service and of the Bureau of Educational Research at the State University of Iowa. In the years between 1932 and 1944, he served as a high school science teacher. He received his B.A. degree in 1932 from the University of Northern Iowa. He holds the M.A. (1936) and Ph.D. (1947) from the University of Iowa.

Ebel is a fellow of the American Psychological Association and a member of the American Association for the Advancement of Science and of the Psychometric Society. He has served as president of the American Educational Research Association, the National Council of Measurements in Education, and of Division 5 of the American Psychological Association. He is author or co-author of several books on measuring educational achievement and of numerous professional articles on the application of educational tests. He edited the fourth edition of the *Encyclopedia of Educational Research*. Ebel is married and has three children and four grandchildren.

Series Editor, Donald W. Robinson

The Uses of Standardized Testing

By Robert L. Ebel

Library of Congress Catalog Card Number: 77-72595
ISBN 0-87367-093-0
Copyright © 1977 by The Phi Delta Kappa Educational Foundation
Bloomington, Indiana

TABLE OF CONTENTS

A Note to the Reader 7

The Task of the Teacher 8

Why Some Teachers Dislike Testing and Grading 9

The Adequacy of Objective Tests 11

How Accurately Can Achievement in Learning Be Determined? 14

Affective Learning and Intangible Outcomes 16

School Achievement and Subsequent Success 17

Beneficial Consequences of Evaluation 18

Harmful Consequences of Evaluation 19

Standardized Tests of Achievement 23

Opposition to Standardized Testing 24

The Case for Standardized Testing 26

What Are Schools For? 27

What Do Standardized Tests Test? 29

How Can Standardized Tests Help? 31

Limitations of Standardized Tests 32

The Charges Against Standardized Testing 33

Knowledge and Behavior 35

The Problem of Validity 36

The Charges of Bias 39

Curricular Influence 40

Formative and Summative Evaluation 42

Criterion-Referenced Tests 44

Alternatives to Testing 48

A Note to the Reader

What I have to present in this fastback is one set of views on the value of standardized tests and on their implications for effective teaching. I hope they are coherent and rational views. But they are obviously not the only views on these topics that competent teachers can hold. For teachers differ. We differ in age, size, sex, and color. We may differ in IQ, though I am uncertain about that because of the indefiniteness of the concept itself. If we do differ in native mental ability, I suspect that the differences are relatively unimportant. Most of us do not begin to use to the limit the genetic potentials for learning with which we are endowed.

But there are other important differences among us. We differ in backgrounds of experience and training. We differ in levels of competence. We differ in our values, concerns, and aspirations. We differ in our beliefs, including our beliefs about evaluation of students and about the nature of effective teaching.

Despite these differences, there are important ways in which we are alike. We are alike in our desire to do a good job of handling our assigned responsibilities, the best job of which we are capable. We are alike in our concern for excellence in ourselves and in our students. Most of us recognize that in the long run there is only one sound basis for satisfaction in the work we do and security in the positions we hold. That basis is our own diligent pursuit of excellence.

My hope is that the views expressed here will stimulate you to reflection and discussion, and, if you are a teacher, that this will contribute to your efforts to teach more effectively.

The Task of the Teacher

Some teachers are not in favor of evaluating student achievement—of giving tests and assigning grades. They may consider themselves to be good teachers. Students may praise them. They may be reputed to be good teachers throughout the school and community. But unless the achievements of their pupils are evaluated, they can provide no solid evidence that they are, in fact, good teachers. Instead they may be merely good entertainers, or good politicians, or good companions. As teachers they may be quite ineffective.

For the essential task of the teacher is to facilitate student learning. That is why schools were established. That is why taxes are collected to maintain them. That is why teachers are employed to staff them. Never underestimate the importance of learning in human society. How well we live depends in large measure on how well we think, and how well we think depends in large measure on what we have learned. The success of every education institution depends on its ability to cultivate student learning.

Now the best way to determine how much learning has occurred is to observe how successfully the student can cope with tasks that require learning. This means testing. It also means grading, for grades can provide concise, meaningful indications of the degree of a student's success in learning. If our testing and grading is educationally unproductive or harmful, it is not because the tools themselves are inappropriate. It is because we lack the skill, or are unwilling to take pains, in using those tools.

Why Some Teachers Dislike Testing and Grading

With such clear, rational justifications for testing and grading, why do some teachers oppose their use? Because the task of using them is difficult and disagreeable. Because some teachers regard the evidence of learning provided by tests and reported by grades to be of limited value, inaccurate, and largely irrelevant to the purposes of education. Because it seems to some teachers that the net effect of student evaluation on the process of education is more harmful than beneficial. Let us take a closer look at these objections to testing and grading.

It would be foolish to claim that constructing a valid test of student learning, or assigning grades validly is a simple, easy task. But the more knowledge of the subject the teacher has, and the more competent he* is in test construction and grade assignment, the easier the task will be and the more satisfactory the outcome. Nowhere is a poor teacher's limited grasp of his subject revealed more clearly than in the triviality, the ambiguity, or the indeterminacy of the test questions he writes. And if the teacher, like most other teachers, has had no systematic, competent instruction in the techniques of testing and grading, should anyone be surprised that he finds the task difficult and the results dissatisfying?

It would also be foolish to expect that a humane teacher could ever enjoy the responsibility of reporting an adverse judgment of the

*For clarity and economy, we use the masculine form of pronouns throughout this publication when no specific gender is implied. While we recognize the trend away from this practice, we see no graceful alternative. We hope the reader will impute no sexist motives; certainly none are intended.
—*The Editors*

quality of another human being's work. But it is a responsibility that a good teacher cannot avoid. For to encourage excellence in learning a teacher must recognize and reward it. This means that the teacher must also recognize and report deficiencies in learning. The more clearly the students understand what they are expected to learn, and why it is important to learn, and the more solidly the teacher's judgments are based on valid evidence of learning, the more acceptable a teacher's evaluations of student learning are likely to be, and the more comfortable the teacher can be in making them.

The Adequacy of Objective Tests

Among teachers who accept the need for evaluating student achievement, and the need to test students as a basis for evaluating their achievements, there are those who believe that the only valid measures of achievement are those obtained from essay testing. They mistrust the value of objective tests and oppose their use. Since most standardized tests, and tests used in wide-scale testing programs are necessarily objective tests, these critics deplore the extensive and increasing use of such tests. Let us consider some of the pros and cons of objective testing.

One of the attractive features of the objective test form is the ease with which it can be scored. But there are others also, such as definiteness in scoring and extensiveness in sampling various aspects of achievement. The scorer of an objective test indicates unequivocally which answer he considered correct to each question. Each student's performance is judged against the same standard. Thus an objective test score is likely to be fairer to the students than an essay test score. Because more questions are asked in the usual objective test than in the usual essay test, a more extensive sample of the student's command of knowledge is obtained. Both fairness in scoring and extensiveness in sampling operate to make objective test scores typically more reliable than essay test scores.

But what of the charges of superficiality, ambiguity, and guesswork that critics lay against objective tests? That test questions, either in objective or essay form, are sometimes trivial and superficial is beyond dispute. But that objective test questions can probe the student's command of essential ideas and his ability to use these ideas to solve complex problems and make important decisions is

also beyond dispute. If an objective test question is trivial, the fault lies less in the form than in the content of the question.

Yet it is surely true that objective test questions appear to be trivial more often than do essay test questions. This is due in part to the fact that an objective test ordinarily includes many more questions than does an essay test. If a test can include only a few questions, the tendency is to make each one general and comprehensive. If it can include many, the tendency is to make each question more narrowly specific.

Objective test questions also tend to be more "factual" than essay test questions because it is easier to defend one particular answer as the correct answer if the question is factual than if it involves inference, explanation, problem solving, or prediction. But it is important to remember that a fact in this sense is a verifiable truth, which need not be trivial. There are many important facts, such as universal gravitation, and $E = mc^2$. Indeed, if a subject for study in school or college is not loaded with important factual truths of these kinds, the value of studying it would seem open to serious question.

Yet another reason why objective test items are sometimes thought to be superficial is that answers to them apparently could be learned by rote, without real understanding. And so they could, but for a number of reasons they seldom are so learned. For one thing, it is always possible to pose questions on an objective test that the examinee has never encountered before, and thus to require answers from him that he could not possibly have learned by rote. For another, rote learning is a difficult, ineffective, and unsatisfying method of learning most things that students study. The student who has difficulty in understanding a subject is unlikely to find rote learning an attractive or effective alternative.

That test questions, either in objective or essay form, are sometimes ambiguous is also beyond dispute. Indeed, the separateness of human beings, the uniqueness of their experiences, and the indefiniteness of language make the total elimination of ambiguity in human communication impossible. Ambiguity hampers our teaching as well as our testing. We need always to seek to avoid it or to minimize it. With reasonable skill and care in test construction it can be reduced to the point where it no longer interferes seriously with the evaluation of learning.

Like ambiguity, guessing is more of a bogey than a genuine

menace in the use of objective tests. Well-motivated students do very little blind guessing on tests that are appropriate for them. The correctness of their informed guesses is related substantially to the amount of relevant information they command. Thus their "guesses" provide valid indications of achievement. A student who does a great deal of blind guessing is likely to get a very low score on a good test. For all these reasons, the influence of guessing on objective test scores is far less than it is commonly assumed to be.

Finally, the effects of both ambiguity and guessing are to make test scores unreliable—that is, to make repeated measurements yield inconsistent results. If a test constructor succeeds in building a test that yields reliable scores, and many objective tests do yield reliable scores, it is safe to conclude that supposed defects related to ambiguity and guessing are not serious in those tests.

Thus despite the criticisms of objective tests, it seems likely that their popularity will continue to grow. The supposed deficiencies of these tests are not inherent in the form, and do not seriously detract from their value. Growing recognition of the fairness and dependability of objective tests, and growing skill in using them will also contribute to their popularity and their effectiveness.

How Accurately Can Achievement in Learning Be Determined?

Consider now the relevance and accuracy of the evidence of student learning provided by tests and reported by grades. The question of accuracy can be dealt with most easily. If a test is composed of a sufficiently large and representative sample of clearly specified tasks of moderate difficulty, highly reliable measures of achievement can be obtained. A reliable test accurately identifies different levels of learning achievement. Inaccuracy in educational measurement is largely the result of inept testing and careless grading.

The problem of relevance is more troublesome, but much of the trouble is of our own creation. In the word of Bishop Berkeley, "We first raise a dust and then complain that we cannot see." The relevance of a good test of achievement in learning is quite easy to defend, for what the test measures is substantially the same as what the teacher has been trying mainly to teach—the command of knowledge. Command implies understanding and ability to use. The knowledge of most use to humans is verbal knowledge, which is to say that it is knowledge which can be expressed, communicated, and stored in words and sentences. "All knowledge," say the philosophers, "is knowledge of propositions," and a proposition is simply a sentence that can be said to be true or false. The particular excellence of human beings among all other forms of life on earth lies in their ability to produce and to use verbal knowledge. Acquisition of this kind of knowledge is the substance of most achievement in learning. How much of it any student has acquired can be determined by giving him appropriate verbal tasks to perform.

If this solution to the problem of relevance seems too simple to be correct, consider that it is simple only in principle. In practice many particular questions arise. What particular knowledge should be learned? What particular tasks should be set to determine how much of it has been learned? Even if this simple outline of the tasks of teaching and evaluating is accepted, there remain sufficient problems and difficulties in applying it to challenge even the best teacher.

Affective Learning and Intangible Outcomes

But is the formulation just presented even approximately correct? Its emphasis is wholly cognitive. What of affective learning? What of intangible outcomes? At the risk of offending advocates of either of these, let me suggest that they be dismissed from serious consideration as major objectives of instruction. Both function better as catchwords in proposals for educational reform than as guides for practical programs for instruction or evaluation.

Intangible outcomes, if they are truly intangible, have no more substance or value than the emperor's new clothes. If outcomes are not truly tangible but simply unanticipated or undefined, they cannot serve as useful foci of instruction or evaluation.

Affective outcomes (that is, values, attitudes, interest, etc.) can be fostered via conditioning (that is, rewards and punishments), modeling (that is, the examples set by other admired human beings), or cognitive learning. The first two of these may be very important in the education of young children, but only the first can be planned purposefully. For achieving affective ends in high schools and colleges, cognitive learning is probably the principal means that should be employed.

A person's affective dispositions inevitably are matters of choice and personal preference. They are more often caught than taught. We ought not try, and probably cannot succeed if we do try, to impose prescribed affective dispositions through purposeful instruction. Also, we probably could not succeed in evaluating any significant part of a student's achievement in learning by assessing his affective dispositions—and probably should not try.

School Achievement and Subsequent Success

Those who question the value of test scores and course grades often cite the low correlations that have been reported between such measures and subsequent success in life. High grades do not infallibly predict success. Low grades do not invariably foretell failure. But if these observations should lead us to conclude that learning has nothing to contribute to living, or that tests cannot measure nor grades report amount of learning, we would be foolish indeed. Evidence that seems to support such unreasonable propositions ought to be examined very closely.

Despite the relevance of learning to successful living, and despite the measurability of achievement of learning, there are at least four reasons why low correlations have been reported. One is that learning is not the only requirement for success in living. It is a necessary but not a sufficient condition. Other factors like ambition, opportunity, personality, and luck have much to do with success.

A second reason is the imperfection of our measures of achievement in learning. Some teachers do not have skill enough, or take pains enough, to measure and report student learning accurately. A third reason is the difficulty of defining success in living and of measuring it reliably. To the extent that our measures of learning and of living are unreliable, the correlation between those measures is bound to be low.

Finally, many of the studies purporting to show low correlations between school learning and subsequent success are small, poorly designed, and badly executed. The sampling errors of moderate to low correlation coefficients are so great that very large-scale studies are required to determine the correlations accurately. If the studies involve only small samples of individuals whose school achievements and vocational successes have been measured crudely, should one be surprised that significant relationships often are not discovered?

Beneficial Consequences of Evaluation

Let us turn now to a consideration of the possible consequences, beneficial or harmful, of the evaluation of student achievement. The beneficial consequences for students are to motivate and give direction to their efforts to learn. If we want excellence of achievement in students we must recognize it and reward it. High test scores and high grades are one means, a very effective means, of providing that recognition and reward. If students know in advance what the tests will require of them, and the basis on which grades will be awarded, the process of evaluation can effectively direct their efforts to learn. I hope you can join me in dismissing as arrant nonsense the assertions that there is something inherently evil in working for recognition or rewards, or that working for high test scores or grades requires *not* working for maximum learning. I personally have never encountered any rational defense of either of those claims.

The beneficial consequence of the evaluation of student achievement for teachers, and for the departments and colleges that employ them, is to provide a basis for justifying continued employment of the teachers—and continued existence of the educational institutions. If teachers and institutions claim to be doing a good job, as most of them do, they should not be surprised or offended if taxpayers and tuition payers ask, "Where is the evidence?" Evaluations of student achievement can provide much of that evidence. So long as public confidence in the soundness and quality of its educational institutions was high, evidence of soundness and quality was seldom demanded. But public confidence has waned, and it will take more than declarations to restore it. And, not incidentally, evaluation of student achievement may do much to help schools and colleges to actually do a better job, and thus to restore public confidence in them.

Harmful Consequences of Evaluation

What of the possibly harmful consequences of using tests and grades in the evaluation of student achievement? What of the anxiety they may generate; of the occasions on which they are said to have caused some pupils to break down and cry? What of the discouraging effects of low test scores; of damage to the pupil's self-concept? What of testing's imposition on teachers of the role of judicial adversary rather than supportive helper? What of the emphasis on competition, and on conformity? What of the temptation for pupils to cheat?

There is, no doubt, anecdotal evidence to support some of these claims as affecting particular pupils. Common sense suggests, however, that the majority of pupils are not harmed by testing. So far as I know there is no substantial survey data that would contradict common sense on this matter. The teachers I talk to seem more often concerned with pupils who don't care enough how well or how poorly they do on such tests than with the relatively rare instances of pupils who seem to care too much.

It is normal and biologically helpful to be somewhat anxious when facing any real test in life, regardless of one's age. But it is also a necessary part of growing up to learn to cope with the kind of tests life inevitably brings. Of the many challenges to a child's peace of mind, caused by such things as angry parents, playground bullies, bad dogs, shots from the doctor, and things that go bump in the night, standardized tests must surely be among the least fearsome for most children. Unwise parental pressure can sometimes elevate anxiety to harmful levels. But usually the child who breaks down in fears at the prospect of a test has problems of security, adjustment, and maturity which testing did not create, and which cannot be solved by eliminating tests.

A pupil who consistently gets low scores on tests covering material which he has tried hard to learn is indeed likely to become discouraged. If this does happen, the school cannot claim to be offering a good educational program, and the teacher cannot claim to be doing a good job of teaching. Most low scores on tests, however, go to pupils who, for one reason or another, have not tried very hard to learn. In the opinion of the teachers of such pupils, it is the trying rather than the testing that is most in need of correction.

That low test scores or low grades are likely to discourage some students cannot be denied. But it will not discourage all of them. Many will accept these indications of incomplete learning as natural and predictable results of their own limited efforts to learn, or of their inadequate backgrounds for learning. For some, the low marks may even be helpful in correcting overly optimistic aspirations. Still, there will be well-qualified students who have worked hard but whose achievements do not justify highest praise. Should teachers protect them from discouragement by refusing to recognize limitations in their achievement? Or does the art of living require all of us to learn to cope with occasional discouragements, taking corrective action if we can, persevering steadfastly if we cannot? Without the possibility of failure, success has no meaning. In the words of Robert Frost,

> Never have I been sad or glad
> That there was such a thing as bad.
> There had to be, I understood,
> In order for there to be good.

Pupils do indeed deserve support from their teachers. But what needs to be supported is the pupil's person and potential, not his mistakes or shortcomings. If teachers support ignorance or wrongdoing, they betray their profession. In the area of school achievement, what teachers need to support is progress toward the goals of learning. Tests can help them to provide that support.

Is competition harmful and undesirable? Surely not always. We rely on competition between manufacturers to give us good products at low prices. We rely on competition between candidates for office, and between political parties, to give us good government. Cooperation, on the other hand, is what is required to keep our families stable and to make our community projects successful.

It is a mistake to regard competition and cooperation as mutually exclusive alternatives in behavior. Each has a contribution to make, an essential role to play. Members of an athletic team cooperate with each other in order to compete successfully with another team. Leaders of factions in a political party compete for the party nomination, and then cooperate to elect the party's nominees.

When competition is destructive, unfair, or inappropriate, it can be evil. But when it leads to excellence in achievement it is surely good. Since excellence is inherently a matter of comparison, the pursuit of excellence inevitably implies some kind of competition. If we were to ban all competition in order to eliminate the possible abuses of it, we would have to give up all hope of excellence as well. The price seems too high. Let us rather concentrate on making competition constructive, fair, and appropriate. The pursuit of excellence in social learning can involve that kind of competition.

Consider next the possible harm student evaluation may do through rewarding conformity. If one is an advocate of nonconformity as a way of life; if one rejects conformity to any popular belief, no matter how soundly based; if one rejects conformity to any conventional mode of behavior, no matter how generally approved and practiced, then one may be concerned justifiably with the kind of conformity that evaluations of student achievement is likely to induce. But if, on the other hand, one regards conformity and nonconformity not as universal, absolute virtues or follies, but rather as conditional goods or evils, depending upon what one is conforming to, then there is little cause for concern. For what students are taught to believe in schools is, on the whole, true, and how they are taught to behave is, on the whole, good.

Fear not that conformity will stifle creativity. Rejection of an old idea is not a precondition for creativity but one of the consequences of it. It is not the first step toward being creative but the last. All of the great creative geniuses of our times, men like Edison, Einstein, Frost, and Picasso, owed their successes not to rejection, but to thorough mastery of the knowledge and skills of the scientists, mathematicians, poets, and artists who went before them. They reached new heights standing on the shoulders of their predecessors.

Does the evaluation of student achievement lead to cheating? Surely evaluation has been the occasion for cheating, but has evaluation caused the cheating? If one is inclined to attribute all personal

errors to social pressures, to blame the whole society for the transgressions of any of its members, then any occasion for a misdeed becomes the cause of that misdeed.

I am not so inclined. The society I live in seems to me to be composed mainly of responsible individuals. They are relatively free to choose to do good or evil, to gain the rewards of doing good, and to suffer the penalties of doing evil. That it is possible for students to cheat, and perhaps to gain an unjustified advantage by cheating, seems to me to offer no acceptable excuse for cheating. To abolish tests and grades in order to prevent student cheating seems no more reasonable than to abolish banks in order to prevent bank robberies. We are less in need of crime-proof institutions, I think, than of crime-proof individuals.

On balance, then, the consequences of evaluating student achievement are far more beneficial than harmful. The supposed harmful consequences may serve as rationalizations for a teacher's failure to do the difficult and sometimes disagreeable job of evaluating student achievement. But they do not excuse or justify that failure.

Standardized Tests of Achievement

Thus far we have been discussing testing and evaluation in fairly general terms. What we have had in mind mainly are the kinds of tests teachers make and use themselves. There is another kind of test, however, that offers other opportunities for assessing pupil achievements and at the same time presents other problems. This is the kind of test produced by specialists outside the local school; perhaps by a commercial test publisher, a university testing service, or a state department of education. Such tests are often referred to as standardized tests. The decision to use them may be made by an individual teacher, but more often the local school administration or the state education agency makes the decision, and requires the use of a particular test or set of tests.

Achievement tests are standardized in at least three ways: 1) in explicit definition of the area of achievement to be sampled by the tasks presented in the test; 2) in exact directions for uniform administration of the test to all examinees; and 3) in the provisions of standards for interpreting individual scores. Often these standards are in the form of tables of norms which report how well the students in some appropriate reference groups succeeded on the test. Thus a test of achievement is standardized with respect to what is tested, how it is tested, and what the test score means.

The content of a standardized test generally reflects a consensus of what is most important to test in an area of learning. Usually such a test is constructed with greater care and expertness than the typical teacher-made test. Efficient though sometimes rather slow machine scoring, reporting, and analysis services are available for many standardized tests of achievement. All of these features have added to their attractiveness and utility. Yet all of them have been the focus of attacks on standardized testing.

Opposition to Standardized Testing

Articles critical of testing appear frequently in the popular press. A number of books dedicated to stating the case against standardized tests have been published. Legislatures in some states have enacted laws prohibiting the use of standardized tests of intelligence. Several professional associations of educators have urged a moratorium on standardized testing until better tests are available and teachers are better trained to use them wisely.

Some of the most vigorous attacks on the use of tests in schools have been directed at intelligence testing. Quite often standardized achievement tests have been lumped together with standardized intelligence tests and tarred with the same brush. There are important differences, I believe. Intelligence tests can be used constructively in schools, but sometimes they have been misused. The scores they yield have sometimes been misinterpreted. A stronger defense can be made for the value of standardized achievement tests than for standardized tests of intelligence in school instruction.

One whose opinions are shaped by these criticisms might conclude that standardized tests have lost favor and are headed for extinction. But the use of standardized tests of educational achievement in the schools of the nation has not declined appreciably. Legislatures in a number of states, including some that have banned intelligence tests, have enacted requirements for additional testing in statewide programs of educational assessment. Tests provide the principal source of data for the National Assessment of Educational Progress, sponsored by the Education Commission of the States. Thus, despite the attacks, standardized testing continues to flourish.

Nevertheless, the attacks are real and substantial. They ought not to be ignored, or dismissed casually, by specialists in educational measurement. If they remain unanswered, they have the potential for considerable harm to the pursuit of excellence in the public schools.

Who are the attackers, and what is the basis of their opposition to testing? The attacks do not come from parents who are concerned with what their children are learning or are not learning in school. They do not come from school board members who are charged with the responsibility for maintaining good local school programs. Opposition to standardized tests of achievement does not come from legislators who want to know how much education their state-aid dollars are buying. It does not come from employers who want workers with well-developed basic skills.

No, the attacks on standardized tests come mainly from certain educators and journalists; from school administrators who fear that they may be blamed, unjustly in their eyes, for substandard pupil average scores on the tests; from teacher organizations like the National Education Association and the National Council of Teachers of English, who fear that their members may be held accountable, unfairly, for low pupil achievements; from spokesmen for minorities who blame the low scores of minority children not on low achievement, but on inappropriate testing; from innovative educators who find emphasis on normal learning in education a handicap to radical innovations in curriculum and instruction; from certain measurement theorists who regard the overthrow of conventional testing practices as a necessary prelude to the popularization of the unconventional alternative they hope to promote. Finally, the attacks come from certain free-lance journalists who stand to profit substantially from the "exposure" of "scandals" in the management of public enterprises, such as education, that are of vital concern to many citizens.

The arguments used in support of these attacks are not necessarily the same as the motivations for making them. The public arguments reflect concern for the inadequacies of the tests and for the potential harm that testing can do to education. If the arguments that are presented could be adequately supported with reason and evidence, they might indeed be sufficient grounds for greatly reducing or totally eliminating standardized testing in the schools. But, on close examination, the arguments often appear to rest on false premises and faulty deductions. Let us therefore examine some of the more prominent of the arguments that have been made against standardized testing in the schools.

The Case for Standardized Testing

First, it may be advisable to set forth the basic rationale for standardized testing. It is this rationale, supported by tests of steadily increasing quality, that is responsible for much of the growth of standardized testing in the last half century. This rationale can be organized in the form of answers to four questions:

1) What are schools for?
2) What do standardized tests test?
3) How can standardized tests help?
4) What are the limitations of standardized tests as educational tools?

What Are Schools For?

If we look for our answer in the expectations of parents, and in the expectations of citizens who pay the bills, schools are for learning. Learning what? In human culture most of what needs to be learned is verbal knowledge. The unique achievement of mankind, the special excellence of the human species among all others in creation is the ability to produce and to use verbal knowledge. The term *verbal* includes here not only words in the ordinary sense, but also the symbols of mathematics, science, logic, etc., which stand for concepts and can be read as words.

Because it can be recorded, stored, and communicated—and above all because it provides the principal food for human thought—verbal knowledge is a very powerful form of knowledge. What pupils spend most of their time trying to do in good schools, and what their teachers spend most of their time trying to help them do, is to gain command of some of the rich store of useful knowledge, expressible in words, that sages and scholars and scientists and explorers and philosophers and poets have placed in trust for them.

But verbal knowledge cannot be conveyed as a gift. Information can be so conveyed, but knowledge cannot. Only when the learner reflects on the information he has received, seeking to understand it, seeking to build it into his own private structure of concepts and relations, is it transformed into knowledge. The student must earn through his own efforts the right to say, "I know." School learning, in short, consists mainly of each student's development of a cognitive structure of knowledge.

Proponents of innovative educational programs sometimes disparage cognitive goals of education, referring to them as "a few facts learned by rote in order to be regurgitated on command of the

teacher." What a grotesque distortion! Facts are only the raw materials. The results we seek are not memorized facts. They are personally posssessed understandings. They are not few in number, but almost limitless. Their function is not to satisfy the teacher, but to increase the competence of the learner, which is no small thing. Cognitive competence is not the only thing that counts toward success and happiness in life, but most of us would be a lot better off if we had a lot more of it. And cognitive competence is the kind of competence the school is best equipped to give to its pupils.

What Do Standardized Tests Test?

Whatever an educational test purports to measure, whether knowledge, or interests, or attitudes, or some trait of personality, the tasks it presents to the examinee will be approached primarily as cognitive tasks if it is a paper and pencil test. These tasks ask, essentially, what do you *know*, about things and ideas in the world, about yourself, about how to get things done. Educational tests are all, in one way or another, tests of cognitive development. But if it is correct for us to say that the job of the school is mainly to facilitate the pupil's cognitive development, then it follows that appropriate educational tests can show how well that job is being done.

What a particular test actually does measure is not always indicated clearly by its title. The most accurate and meaningful designation is provided by a description of the kind of tasks it presents, not by a name for the trait or characteristic it was intended to measure. If a test is called a dictation spelling test, or an error recognition spelling test, or a proofreading spelling test, one knows more exactly what is being tested than if it is called simply a spelling test. How well the scores on these three types of spelling tests may correlate in any particular group of pupils is an empirical question. Depending on how the pupils have been taught, the correlations might be higher or lower. One cannot assume that because they all involve spelling in some way they are all tests of the same educational achievement. Many of our problems of test validity or bias would never arise if we described what our tests measure in terms of the domain of tasks they sample instead of in terms of the more generalized, but less clearly defined, traits they were intended to measure.

In sum, standardized tests of educational achievement measure

the structure of knowledge a pupil has developed in some particular, defined area of school learning. Since helping pupils to develop such cognitive structures is a major part of the job of the schools, then standardized tests can make a major contribution toward getting the job done.

How Can Standardized Tests Help?

Standardized tests can help by keeping the eyes of all concerned with the learning process fixed on the main target. By revealing variation in the effectiveness of different instructional procedures or different curricular arrangements. By motivating teachers' efforts to teach well and pupils' efforts to learn well. By recognizing and rewarding success in learning. By making possible comparisons—between pupils, between teachers, and between schools—of the outcome of efforts to learn. By causing schooling to become a purposeful educational enterprise whose results can be assessed systematically. No school can do a good job and show that it is doing a good job without systematically auditing the results it is getting. Standardized tests of educational achievement provide one good means for making such audits.

Limitations of Standardized Tests

There are two major limitations of standardized achievement tests. One is that good tests are likely to be available only in general areas of learning. Production, distribution, revision, and general maintenance of tests in special subject areas tend to cost more than low volumes of sales can support. But where standardized testing is used to make general surveys of pupil and school achievement, the lack of good tests in specialized subject areas is no serious handicap.

The other major limitation results from technical imperfections in the tests themselves. The achievement to be sampled by the test is not always precisely defined. The process of sampling may be more haphazard than systematic. The item writing may be imperfect, involving ambiguous tasks, questionable answers, or giveaway clues. Finally, the norms may be too limited or too biased to reflect an examinee's level of achievement accurately. More expertness and care in test construction can minimize this second type of limitation. Hence, while standardized tests of school achievement do have limitations, they do not seriously reduce the value of such tests in education.

The Charges Against Standardized Testing

Having tried to make something of a case for standardized testing in the schools, let us consider some of the specific attacks that have been made on it. One group of arguments focuses on the limited information provided by a standardized test. It measures only the pupil's knowledge, say the critics, ignoring more important characteristics such as responsibility, initiative, and originality. Further, the critics argue, the knowledge measured is at the lowest level, recognition or simple recall, and requires none of the higher mental processes.

The fundamental importance of knowledge in human affairs has already been stressed in this paper. Perhaps two other points, one dealing with the levels of mental process, and the other with the relation of knowledge to behavior, deserve comment here, in relation to the "limited information" charge.

It is simply not true that standardized tests are made up of items requiring *only* recognition or simple recall. Of course recognition of words and symbols and recall of relevant facts and principles are involved in answering all of the questions. The student needs this information to solve the verbal or numerical problem presented in the question. But ordinarily he needs much more. Only if an identical question were presented during the course of instruction could he recognize or recall the answer directly. So numerous are the different questions that may be asked in any area of study, and so interrelated and overlapping, that there is no need to ask the same question on a test as was asked in teaching, and there is only a very small probability that duplication might occur by accident. While it may appear that many items *could* be answered on the basis of recognition or recall, as indeed any item on any paper and pencil test

could be answered, few if any items are likely to be answered solely or primarily on that basis.

To speak of levels of mental process is necessarily to speak vaguely. We know very little about mental process as a whole, to say nothing about different levels in those processes. That some problems are more complex, more difficult to solve, than others seems obvious. It is more difficult to add two common fractions than to add two single digit whole numbers. More difficult than both is the inversion of a matrix. One must know more to solve complex problems than to solve simple problems. But who can say that the mental processes are "higher" in one case than in the other. It seems reasonable to suppose that the main difference is that complex problems require longer sequences of thought processes, whatever they may be. At this time there is no good evidence that different levels of mental processes actually exist. Until that evidence appears, the charge that standardized tests neglect the higher mental processes cannot be taken seriously.

Knowledge and Behavior

Consider now the relation of knowledge to behavior. If behavior is rational it must be based on knowledge: knowledge of purposes, of means, and of consequences. But knowing is not the same as behaving. Knowing how to do something does not guarantee that it will be done. There is also the matter of willing oneself to do it. Knowing it is a necessary, but not a sufficient, condition for effective doing.

Few would deny that good schools can use systematic instruction to help willing pupils acquire knowledge. But what about acquiring the willingness? What about the responsibility, the initiative, the originality? Can these be developed by systematic instruction? So far as I know they cannot. They can be encouraged, recognized, and rewarded, but they cannot be imparted directly. Nor are there tests, paper and pencil, performance, or any other kind, that will measure satisfactorily such traits of typical behavior.

To sum up, if standardized tests provide good measures of a pupil's command of useful verbal knowledge, they do a very important job. They are not limited to recognition and recall. They need not be concerned with allegedly higher mental processes. If they do not measure behavioral traits that schools cannot teach systematically, and that no test can assess adequately, should they be faulted for that?

The Problem of Validity

A second focus of attack on standardized testing is the validity of the tests. This is often regarded as a very serious charge, the answer to which may be complicated, extensive, unconvincing, and basically inadequate. When a measurement specialist sets out to discuss validity in general, or the validity of a specific test in particular, many in his audience are likely to lose both understanding and interest long before he finishes. And it may be, where achievement tests are concerned, that many of the complexities of the validity problem are of our own creation, complexities that we can avoid should we choose to.

The validity of a test has to do with the extent to which the test measures what it is intended to measure. If we limit our claim for what the test was intended to measure to a description of the kinds of tasks composing it, then the test is valid to the extent: 1) that the description is accurate, and 2) that the test yields reliable scores. No criterion measures or validity coefficients are involved. This is direct, primary validity. It involves an operational definition of what is being measured, and is the kind of validity almost all physical measures have. It is the best kind for a test to have.

What complicates the validity problem for some is their interest in measuring, not directly observable task proficiencies, but the strengths of hypothetical underlying traits which presumably are responsible for the proficiencies. The tests we use are often named for the traits they are intended to measure, rather than for the kind of tasks they present. If a test requires the solution of everyday problems it is not likely to be called a test of everyday problems, but a test of practical judgment. If a test requires the examinee to offer suggestions, it is more likely to be called a creativity test than a suggestions

test. If a test offers verbal, numerical, and symbolic problems it is likely to be called an intelligence test, not a problems test.

With tests of hypothetical trait constructs, questions about validity do arise, and often they are very difficult to answer. The problem is not so much to show *that* they measure what they are intended to measure as it is of defining clearly *what* they were intended to measure. Without a clear definition of the intention, validation is impossible in principle. With only a loose, fuzzy conception of what the test is intended to measure, the process of validation is correspondingly loose and fuzzy, and generally yields unsatisfactory results. The more clearly precise the definition, the more straightforward the process and the more conclusive the results.

For standardized tests of achievement, it is usually quite easy to define what the test is intended to measure; that is, how well the pupil can succeed in doing tasks of a specified type. The final test for my course in classroom testing, for example, is intended to measure how well the students enrolled can select the true statements, in a list of 100 or so, related to ideas presented in the text and discussed in the class. If the 100 should include any statements unrelated to ideas presented and discussed, if there is bias in sampling these statements from the domain of possible statements, or if the test does not yield reliable scores, then the test is to that degree invalid.

Standardized tests of achievement, in short, are validated by inspection, with the help of a definition of what the test is intended to measure. Inspection is the means used to validate a diamond, the books of a business, or the title to a property. Why some test specialists insist it is inappropriate to validate an achievement by inspection is hard to understand. Their views on this issue probably should be rejected. If they are, and if standardized tests of achievement are validated by inspection, many of them will be found to have quite satisfactory validity as measures of the specific achievement their tasks permit them to measure.

Note well that this process of test validation calls for a rather specific definition of what the test is intended to measure. The kind of tasks which will make up the test, and the domain from which they will be sampled, must be clearly defined. A spelling test will not be just a spelling test. It will be, for example, a test of words sampled from a particular list, dictated by the teacher and written by the pupils. A particular reading test may be a test based on a specified

type of reading material, read for a particular purpose, and with comprehension tested using a certain type of task.

The particular validity approach greatly simplifies the problem of validating achievement tests. It relieves the test constructor of the practically impossible task of showing that one particular test of spelling, or reading, or arithmetic is a valid measure of some hypothetical general ability to spell, to read, or to do arithmetic. But it requires acceptance of the view that school achievements are also particular, involving acquisition and gradual integration of a vast and diverse assortment of particulars. It requires abandonment of the notion that what we are trying to develop by schooling and to measure with our tests are some underlying general abilities. Not all educators and test specialists may be willing to follow this path. Those who are not, it seems to me, are faced with a very difficult task in trying to show that the achievement tests they make or use are valid. How can you show that a test measures what it was intended to measure if you don't know at all precisely what it was intended to measure?

The Charges of Bias

Standardized tests of educational achievement have also been attacked for their alleged bias against cultural minorities, and against pupils with poor reading skills. The reason for the attack, at least in part, is the fact that such pupils tend to make lower scores on the standardized tests. But surely lower scores alone do not signify bias. If they did, every spelling test would be biased against poor spellers, and every typing test against persons who never learned to type.

How could evidence of bias in a test of educational achievement be produced? To get it one would need a set of biased tests and a set of unbiased tests of the same achievement. Then if a person or group made consistently lower or higher scores on the biased than on the unbiased tests we would have evidence of bias. But note the problem. If they are tests of the same achievement they must be composed of items meeting the same task specifications, and hence drawn from the same item pool. Yet if one set is biased and the other unbiased, there must be some systematic difference between them. To show bias we must show the tests to be the same and yet to be different. This is obviously a logical impossibility. For this reason demonstration of bias in a properly validated achievement test is impossible, and if bias cannot be demonstrated, there is no good reason to believe that it exists.

That a pupil who does poorly on a particular test in English might do better if the test were in Spanish, or if the questions were presented orally, does not mean that the original test is biased against them. It simply means that they do not have much of the precise achievement that particular test measures. Its linguistic context is part of the test. The particularity of what the test measures does not constitute bias.

Curricular Influence

Another direction of attack on standardized tests of achievement relates to their effect on the curricula of local schools. It is charged that they enforce curricular conformity; that they fail to test what local schools or particular teachers have been trying to teach; that they hamper curricular innovations or the use of open curricula. There is substance to these charges, but the effects are neither so overpowering nor so harmful as the critics imply.

If a school wishes its pupils to do well on certain tests, it must see to it that pupils receive instruction in the achievements sampled by the tests. But these are areas that panels of expert teachers have identified as important for most pupils to learn. A teacher or school that decides to concentrate on achievements other than these should be prepared to do two things: 1) to argue convincingly to the public that what they are trying to teach is indeed more important for these pupils to know than the things the standardized test is testing; and 2) to prepare, give, validate, and report the results on a local test that does measure what they have been trying to teach. Few of those who complain about the curricular restrictions of standardized testing are prepared to do either of these things.

A school that is teaching what the tests test will surely teach many other things beside. Even in the basic skill or core areas that the test does sample, there will be class time and teacher time to venture into interesting and useful areas of learning not covered by the standardized tests. Standardized tests can dominate local curricula only to the extent that school administrators and school teachers allow them to.

In some programs of open education there are covert goals for learning, and classroom activities are manipulated so that pupils

make systematic progress toward these goals. In such programs, standardized tests of achievement may be not only appropriate but essential. Other programs have no particular goals for pupil learning and are satisfied merely with maximum pupil freedom, trusting nature alone to do the job that others employ the art and science of teaching to help nature to do. That standardized tests of achievement are inappropriate for such programs may imply more criticism of the programs than of the tests.

Formative and Summative Evaluation

Tests given after instruction and study to assess how much has been learned are said to provide data for summative evaluation. Tests given during or before instruction for the purpose of helping the teacher to teach or the pupil to learn are said to provide data for formative evaluation. Some say that formative evaluation is more valuable and less objectionable than summative evaluation. They suggest that teachers should be more concerned with the formative than with the summative.

Several arguments are advanced to support this suggestion. The teacher's job is to help pupils to learn, not to stand in judgment over them. If pupils have not learned, the process of instruction is probably at fault. The goal of formative evaluation is to improve that process. Summative evaluation reveals shortcomings when it is too late to do anything about them.

There is some force to each of these arguments, but none of them, independently or taken together, accounts for the popularity of formative evaluation in some quarters. That popularity derives from the presentation of formative evaluation as an alternative to summative evaluation. If summative evaluation could be minimized or eliminated altogether, pressure on pupils and teachers for achievement in learning would decline or disappear. The threat of accountability would fade.

But formative evaluation is really no substitute for summative evaluation. Each has an important role to play in the educational process. Without summative evaluation of some kind we would not know whether the formative evaluation had done any good. Without formative evaluation of some kind, instruction is likely to falter, and make the results of summative evaluation disappointing.

True, the teacher's job is to help pupils learn. But one of the ways of helping them learn is to set goals for learning, to recognize and reward achievement in learning. This calls for summative evaluation. If pupils fail to learn, is it the fault of the process of instruction, a fault that formative evaluation might correct? It might be the fault of instruction, but then it might also be the fault of the student, his weak background, or his lack of interest and effort. Learning is a cooperative enterprise. To assume that total control of the enterprise, and total responsibility for its success rests with the teacher is to substitute fantasy for realism.

If failure to learn is due to ineffective instruction, could formative evaluation correct the fault? The probability is not great. Ineffective instruction is usually the result of incompetence or indifference on the part of the teacher. Formative evaluation cannot correct either incompetence or indifference. It can help a capable, conscientious teacher work toward increasingly effective teaching, but the progress is likely to be gradual. Formative evaluation cannot be counted on to suddenly transform an ineffective instructional process into an effective one.

Consider finally the argument that summative evaluation reveals shortcomings when it is too late to do anything about them. Again, the argument has some substance, but its weight is easy to overestimate, and often has been overestimated. It is never too late to do something about one's shortcomings until life itself is ended. After that, the shortcomings no longer matter. If the first indication of a student's shortcomings in learning comes with a final examination, either the student is singularly unperceptive, or his teacher is singularly indifferent to the progress in learning that his students are making.

Any evaluation has both formative and summative contributions to make. The verbal difference between formative and summative evaluation is much sharper and more clear-cut than their actual differences in form or function. To replace summative with formative evaluation is not only unnecessary and unwise; it is logically impossible.

Criterion-Referenced Tests

Until quite recently the existence of norms for interpreting scores on standardized tests has been regarded as a special value of such tests. Now a different kind of testing, called criterion-referenced, or objectives-referenced, is being advocated. Among the advocates are some very influential measurement specialists. If they claimed only that objectives-referenced tests deserve a place alongside norm-referenced tests in a school's testing program there would be no reason to challenge the claim. But when they demand that schools throw out their standardized tests and replace them with objectives-referenced tests, when they blame standardized testing for the shortcomings of contemporary school programs, one ought to look closely at their logic and evidence.

It is possible to give both an objectives-referenced and norm-referenced interpretation to the number of correct answers a pupil gives to the questions on any achievement test. But an objectives-referenced test, as conceived by many of its advocates, differs in at least two important respects from a typical norm-referenced test: 1) in the degree to which items are focused on specific learning objectives, and 2) in the way in which correct responses are aggregated to obtain a test score.

The items in a norm-referenced test sample the area of achievement independently of each other and representatively of the whole domain. The items in an objectives-referenced test, on the other hand, focus in small clusters on the attainment of a limited number of specific objectives of learning in the area. The score on a norm-referenced test is usually based on the total number of correct responses given. The score on an objectives-referenced test, if a total score is obtained at all, is usually the number of objectives achieved.

If the pupil answers a sufficient proportion, say three out of five or seven out of 10, of the items on a particular objective correctly, the pupil is said to have achieved that objective.

Advocates of objectives-referenced tests claim other important characteristics that differentiate them from norm-referenced tests. It is said that the type of item required to tell whether or not an objective has been achieved is different from the type required to tell how much a pupil has learned in an area of study. No examples of items which demonstrate this difference have come to my attention, and no closely reasoned arguments have been offered to show why it ought to be true.

It is said also that conventional item and test statistics are inappropriate for judging the quality of an objectives-referenced test. The basis for this claim is the assumption that item and objective scores on an objectives-referenced test typically show little or no variability. While this situation is a theoretical possibility, evidence that it is a frequent reality is seldom, if ever, offered. Indeed the scores on objectives-referenced tests used in the Michigan State Assessment Program tend to show extreme variability, associated with U-shaped score distributions.

Further, if lack of variability does occur in practice, it must be due to the fact that everyone, or no one, in the tryout population could answer the item correctly. The remedy for this is obviously a more heterogeneous tryout population, including both those who have and those who have not attained the objective. The job of any test, whether norm- or objective-referenced, is to discriminate levels of achievement, high or low; to discriminate mastery from lack of mastery. Given appropriate tryout populations, conventional item and test statistics are perfectly capable of doing that job.

If the principal distinguishing characteristics of objective-referenced tests are, 1) subsets of clusters of items focused on particular objectives of learning, and 2) pass-fail scoring of pupil responses to items in each objective cluster, do objectives-referenced tests have any general advantage or disadvantage when compared with norm-referenced tests? It goes almost without saying that both kinds of tests should be, and usually are, developed to measure the extent of pupil attainment of designated objectives of instruction. The principal differences between them are in the detail and specificity with which the objectives are stated, and in the

degree of concentration of items on particular objectives. An objective-referenced test of 100 items is likely to consist of five-item clusters measuring attainment of each of 20 specifically stated objectives. The items within each cluster will usually be closely alike. A norm-referenced test of 100 items, on the other hand, is likely to consist of 100 separate, distinct items, all related to the general objectives of instruction in an area of learning, but not tied to particular specifically stated objectives.

In areas of learning where objectives of instruction are relatively few, clearly distinct from each other, easy to specify exactly, and relatively difficult to attain, specifically stated objectives can be very useful. But most areas of school learning do not have these characteristics. For such areas, a specific statement of numerous detailed objectives is likely to be tedious and arbitrary, without substantial educational rewards. There are also some dangers to objectives-referenced testing in areas of learning where the objectives are not few, distinct, easy to specify, or hard to believe. It may make teachers stress recall of discrete elements of information instead of encouraging pupils to develop integrated structures of understanding. Teachers may be reluctant to allow pupils to consider any aspects of the subject, however interesting or relevant, that are not specified as objectives. It seems to be the case that in many areas of learning, good teaching can be hampered by objectives that are too specific as well as by objectives that are not specific enough.

Use of objectives-referenced tests involves another problem—that of setting a minimum score that indicates achievement of an objective. This often involves arbitrary judgments that are hard to justify. Is three out of five items answered correctly on one objective enough? Why not four out of five or five out of five? Further, if a teacher is interested in the most accurate measures of pupil levels of achievement, counting the number of objectives achieved is almost certainly not as good as counting the number of items correctly answered.

Before leaving the question of norm- versus objectives-referenced tests, it may be appropriate to say a word or two in defense of norms. For years test publishers have worried about the quality of the norms for their standardized tests. Are they based on sufficient samples? Are the samples adequately representative? Are

they differentiated so as to meet the needs of particular users? Were they obtained from testings comparable to regular operational testings? Hence it comes as something of a shock to be told that norms are really unnecessary after all; that the important thing is what a pupil knows or can do, not how his knowledge or ability compares with that of his peers; that a school should judge its accomplishments by its own standards, paying no attention to the accomplishments of other schools.

Those who are interested in excellent education find it difficult to believe that comparisons of educational achievement are irrelevant or unnecessary. No proud parent believes it. No capable teacher believes it. No serious student believes it. It simply is not true. Only the promoters of objectives-referenced tests, who save themselves great trouble and much money by not providing norms, have tried to tell the world that their deficiency is in fact a virtue.

But clearly it is not. The only basis for judging a human performance excellent, acceptable, or inferior is in comparison with other human performances. The only basis for setting reasonably attainable objectives for pupil learning is knowledge of what similar pupils have been able to learn. Of course it is good for a pupil to compare his present performance with his own past performance. But that is no substitute for comparing his present performance with the present performances of his classmates. Surely it is cold comfort to a pupil when his arithmetic teacher says, "For you, five out of 10 is quite good."

True, not all pupils can achieve excellence in learning, no matter how hard they may try. But part of growing up is to find out about one's limitations, to change them if possible, to accept them if not. There are many kinds of excellence in human activities, and they are not all highly correlated. "Get up something to say for yourself," said Robert Frost, by which he probably meant, "Find something worth doing that you can do well, and work to learn how to do it even better." Surely getting rid of comparisons, trying to abolish excellence in the hope that deficiency will also disappear, is not the answer.

Alternatives to Testing

Those who disapprove of standardized tests sometimes suggest alternative bases for evaluating the process and products of learning. Among these are direct observations of pupil behavior; collections of the written work pupils produce; interviews with pupils, parents, and teachers; and locally oriented criterion-referenced tests. Systematic efforts along these lines, it is said, would reduce the need for, and the dependence on, standard test scores.

Perhaps it might, to some degree. But are we really faced with the problem of choosing between standardized tests and these suggested alternatives? Mightn't we rather welcome the availability of all bases for evaluation, and use each to the extent that it gives a useful return for the effort expanded? Surely a wise teacher, seeking to do a good job of evaluating pupil achievement, will want to use any productive means that is available. That is what most teachers do at present, I believe, to the extent that they have the time, the ability, and the desire to do so.

The proposed alternatives have some fairly serious drawbacks. They are time-consuming to use. They tend to be subject to substantial errors of sampling, of subjective judgment, of favoritism and bias. Tested against the usual standards of reliability and validity that should characterize any basis for evaluation (and they are rarely, if ever, so tested), they are likely to show up badly. Thus while they may be considered alternatives, they should not be regarded as equally good or superior substitutes. The job they do is different from the job standardized tests do. They are alternatives only in the sense that a taxi is an alternative to an airplane.

Finally, to the extent that these alternatives are used as a basis of honest evaluations, instead of a means of escape from evaluation,

they are subject to most of the criticisms that have been leveled at tests. They may stimulate anxiety. They may discourage. They require the teacher to act as a judge. They encourage competition and reward conformity. They may lead to dissimulation and deception. The fact of the matter is that most criticisms of tests are actually criticisms of evaluation itself. But few would suggest that education can get along without evaluation. It would be absurd to argue against evaluation while arguing for excellence in education. Without evaluation how can one distinguish excellence from mediocrity? So let us continue to use classroom tests and standardized tests along with any other valid indications of achievement in learning that can be used efficiently.

This book and others in the series are made available at low cost through the contributions of the Phi Delta Kappa Educational Foundation, established in 1966 with a bequest by George H. Reavis. The foundation exists to promote a better understanding of the nature of the educative process and the relation of education to human welfare. It operates by subsidizing authors to write booklets and monographs in nontechnical language so that beginning teachers and the public generally may gain a better understanding of educational problems.

The foundation exists through the generosity of George Reavis and others who have contributed. To accomplish the goals envisaged by the founder, the foundation needs to enlarge its endowment by several million dollars. Contributions to the endowment should be addressed to the Educational Foundation, Phi Delta Kappa, Eighth and Union, Box 789, Bloomington, Indiana 47401. The Ohio State University serves as trustee for the Educational Foundation.

You, the reader, can help us improve the PDK foundation publications program. We invite you to comment on the strengths and weaknesses of this fastback. Let us know what topics you would like us to deal with in future fastbacks. Address Director of Publications, Phi Delta Kappa, Eighth and Union, Box 789, Bloomington, Indiana 47401.

All 94 titles can be purchased for $33 ($27.50 for Phi Delta Kappa members).

Any six titles $4 ($3 for members); any eight titles $5 ($4 for members). Single copies of fastbacks are 75¢ (60¢ for members).

Other quantity discounts for any titles or combination of titles are:

Number of copies	Nonmember price	Member price
10- 24	48¢/copy	45¢/copy
25- 99	45¢/copy	42¢/copy
100-499	42¢/copy	39¢/copy
500-999	39¢/copy	36¢/copy
1,000 or more	36¢/copy	33¢/copy

These prices apply during 1977. After that, they are subject to change.

Payment must accompany all orders for less than $5. If it does not, $1 will be charged for handling. Indiana residents add 4% sales tax.

Order from PHI DELTA KAPPA, Eighth and Union, Box 789, Bloomington, Indiana 47401.

The fastback titles now available are:

1. Schools Without Property Taxes: Hope or Illusion?
2. The Best Kept Secret of the Past 5,000 Years: Women Are Ready for Leadership in Education
3. Open Education: Promise and Problems
4. Performance Contracting: Who Profits Most?
5. Too Many Teachers: Fact or Fiction?
6. How Schools Can Apply Systems Analysis
7. Busing: A Moral Issue
8. Discipline or Disaster?
9. Learning Systems for the Future
10. Who Should Go to College?
11. Alternative Schools in Action
12. What Do Students Really Want?
13. What Should the Schools Teach?
14. How to Achieve Accountability in the Public Schools
15. Needed: A New Kind of Teacher
16. Information Sources and Services in Education
17. Systematic Thinking About Education
18. Selecting Children's Reading
19. Sex Differences in Learning to Read
20. Is Creativity Teachable?
21. Teachers and Politics
22. The Middle School: Whence? What? Whither?
23. Publish: Don't Perish
24. Education for a New Society
25. The Crisis in Education is Outside the Classroom
26. The Teacher and the Drug Scene
27. The Liveliest Seminar in Town
28. Education for a Global Society
29. Can Intelligence Be Taught?
30. How to Recognize a Good School
31. In Between: The Adolescent's Struggle for Independence
32. Effective Teaching in the Desegregated School
33. The Art of Followership (What Happened to the Indians?)
34. Leaders Live With Crises
35. Marshalling Community Leadership to Support the Public Schools
36. Preparing Educational Leaders: New Challenges and New Perspectives
37. General Education: The Search for a Rationale
38. The Humane Leader
39. Parliamentary Procedure: Tool of Leadership
40. Aphorisms on Education
41. Metrication, American Style
42. Optional Alternative Public Schools
43. Motivation and Learning in School
44. Informal Learning
45. Learning Without a Teacher
46. Violence in the Schools: Causes and Remedies
47. The School's Responsibility for Sex Education
48. Three Views of Competency-Based Teacher Education: I Theory
49. Three Views of Competency-Based Teacher Education: II University of Houston
50. Three Views of Competency-Based Teacher Education: III University of Nebraska
51. A University for the World: The United Nations Plan
52. Oikos, the Environment and Education
53. Transpersonal Psychology in Education
54. Simulation Games for the Classroom
55. School Volunteers: Who Needs Them?
56. Equity in School Financing: Full State Funding
57. Equity in School Financing: District Power Equalizing
58. The Computer in the School
59. The Legal Rights of Students
60. The Word Game: Improving Communications
61. Planning the Rest of Your Life
62. The People and Their Schools: Community Participation
63. The Battle of the Books: Kanawha County
64. The Community as Textbook
65. Students Teach Students
66. The Pros and Cons of Ability Grouping
67. A Conservative Alternative School: The A+ School in Cupertino
68. How Much Are Our Young People Learning? The Story of the National Assessment
69. Diversity in Higher Education: Reform in the Colleges
70. Dramatics in the Classroom: Making Lessons Come Alive
71. Teacher Centers and Inservice Education
72. Alternatives to Growth: Education for a Stable Society
73. Thomas Jefferson and the Education of a New Nation
74. Three Early Champions of Education: Benjamin Franklin, Benjamin Rush, and Noah Webster
75. A History of Compulsory Education Laws
76. The American Teacher: 1776-1976
77. The Urban School Superintendency: A Century and a Half of Change
78. Private Schools: From the Puritans to the Present
79. The People and Their Schools
80. Schools of the Past: A Treasury of Photographs
81. Sexism: New Issue in American Education
82. Computers in the Curriculum
83. The Legal Rights of Teachers
84. Learning in Two Languages
84S. Learning in Two Languages (Spanish edition)
85. Getting It All Together: Confluent Education
86. Silent Language in the Classroom
87. Multiethnic Education: Practices and Promises
88. How a School Board Operates
89. What Can We Learn from the Schools of China?
90. Education in South Africa
91. What I've Learned About Values Education
92. The Abuses of Standardized Testing
93. The Uses of Standardized Testing
94. What the People Think About Their Schools: Gallup's Findings

See inside back cover for prices.